

Designing A Spell Checker in Indian Language Paradigm: Oriya – A Case Study

Sanghamitra Mohanty

Khyamanidhi Sahoo

Prabhat Kumar Santi

sangham1@rediffmail.com

khyamanidhi@yahoo.com

pksanti@rediffmail.com

RC-ILTS-ORIYA

Dept of Computer Science & Application,

Utkal University, Bhubaneswar,

Orissa, India - 751004

Key words: Run-time data structure, corpus, non-words, real-word error, error pattern, modifiers, conjuncts and optimal suggestion list.

Abstract

This paper focuses on different aspects of designing spell checkers in Indian languages and a systematic approach to the design of an Oriya language spell checker is presented as a case study. Besides error patterns like substitution, transposition, insertion and deletion, emphasis is given on the modifiers and their positions with respect to the consonants and conjuncts being modified. The system also takes care of phonetic errors taking the help of phonetic character hash table. The algorithm generates suggestion set for misspelled words by forward and backward matching with the dictionary words and minimum edit distance classifiers, considering the positional modifiers. Also the system implements a run-time tree structure for dictionary words, according to their starting character and length, for efficient processing.